

TTATAGTCGCTTTGTTAAATTGGCCCTCGAGGTCGACCCACGGCTCGGGCTGCTTGTCCCGTTCCCTGTGCCTGATGTCTATC

Figure 1-A

*T. reesei* *bgI4* nucleotide sequence (non-coding is in bold, coding is in regular font)

TGCCGTTGGCCCTCCTCATCTCCTGTTGTCTGTCTCCTTAGTGCTTCAGTGACGCTAGGTTGGTCACTTTGT  
CCCCCTTCGTTGCTCTGGTGTGCCAAGGTCACCTGCAAGTGGTTTGAACCTTGATATCCTGCTTGAGCATCCGCGT  
CGCCATATAGAGCAGCATATTCTTCTATCTCCTCAAGATCCCCCTCACCGAGAGTTCTATTACCCCGACCCCTTGCTGTCA  
CCAGTCCCTTCCATCATGGCTGATATTGATGTTGAGGCCATCTTGAAGAAGCTCACCTGGCCGAGAAAGTCTGATCTGCTGG  
CTGGTATCGACTTCTTGGCACACAAAGGCTCTCCCCAAGCATGGAGTCCCTCTCTCCGCTTTACAGATGGCCCCCAAACGGCG  
TAAGAGGGACCAAGTTCTTCAATGGCGTCCCTGGGCTGCTTCCCTTGGGACACGTCGCTCGGTTCCACATTCAACCAAAA  
CTCTGCTCGAAGAGGCAGGTAAAGATGATGGGCAAGAGGCCATCGCTAAGAGTGGCATGTGATCCTCGGCCCGACTATCA  
ACATGCAACGCTCCCTCTCGGTGGAAGTGGCTTCGAGTCGATTGGTGAGGATCCGTTCTTGGCGGCTTGGGAGCTGCGG  
CTCTCATCCGCGGCATTAGAGCACTGGAGTGCAGGCTACGATCAAGCACTTTTGTGCAATGATCAGGAGGACAGGCGCA  
TGATGTCAGAGCATCGTCAAGGAGCGGCTCTCCGTGAAATCTACGCACCTCCGCTTCCAGATTGCTGTGCGAGACTCCC  
AGCCGGGTGCGTTTATGACGGCGTACAAATGGCATCAATGGCGTGTCTGTCAGCGGAGAACCTTAAATATCTTGATGGGATGC  
TTCGAAAGGAATGGGTTGGGATGGCTAATCATGAGCGACTGTTACGGCACATACAGTACCAAGAACCGCTTGTGGCAG  
GCCTCGACCTCGAGATGCCCGACCTCCACGCTTCCGAGGAGAAACACTCAAGTTCAAAGTCTCCAAACGGAAAGCCCTTTA  
TCCAGTCAATTGACCCAGAGGCTAGGGAAGTTCTTCAGTTCTGTCAGAAAGTGTGCTGCTCCGAGTGAACGAGAACCGGCC  
CCGAGACGACTGTCAACAAACACCCCGAAACGGCAGCTCTCCTCCGGAAGTTGGCAACGAGGGCATCGTGTGCTGAAAGA  
ACGAGAAACAAAGTTCTGCTGCTTGGCAAGAAAGAACGCTGATTTGTCGGCCCAAACGCCAAGCAGGCCACATACCAAG  
GCGGAGGCTCTGCGGCACTCAGGGCTACTACGAGTCACTCCCTTTGACGGCTCAGCAAGCAGCTCGAGACGCCGCCAT  
CGTACACCGTCGGCGCTACACCAACCGTTCTCCCATTTAGGCGAGCAGTGCCTCACGCCGACGGCGCTCCGGGCATGC  
GCTGAGGCTCTTCAACGAGCCCCCTGGTACCCCTAACCGCCAGCACATTGACGAGCTCTTCTTACCAAGACGGAATGC  
ACCTGGTGGACTACTACCAACCCCAAGCGGCAGACACGTGGTACGCCGACATGGAGGGCACGTACACCGCCGACGAGGACT  
GCACCTACGAGCTCGGCTCGTCTGCGGCAACGGCAAGGCGTACGTAGACGACCAAGTCTGTCGTGACAAACGCCACCA  
AGCAGGTCCTCCGGGATGCTTCTTGGCTCCGCCACCCGCGAGGAGACGGGCCGCTCAATCTCGTCAAGGGCAACACGT  
ACAAATTCAAGATCGAGTTCGGCTCCGACCCCACTACACCTCAAGGGCGACACCATCGTCCCGGCCACGGCTCCCTCC  
GCGTCGGCGGCTGCAAGGTCAATTGACGACACGAGCCGAAATCGAAAAGTCCGTGCGCTCGCCAAAGGAGCACGACGAGTCA  
TCATCTGCGCGGGCTTAAACGCCGACTGGGAGACCGAGGGCGCCGACCGCGGAGCATGAAGCTCCCGGGCGTGTGGACC

Figure 1-B

*T. reesei* *bgl4* nucleotide sequence (non-coding is in bold, coding is in regular font)

AGCTCATTGCCGACGTGGCCGCGGAACCAACACCGTCGTCATGCAGACGGGCACCCCGAGGAGATGCCCTGGC  
TCGACGCCACGCCCGCGTCATCCAGGCCTGGTACGGCGGCAACGAGACGGGCAACTCCATTGCCGACGTCGTCTTTGGCG  
ACTACAAACCCCTCGGGCAAGCTGTCCCTCAGCTTCCCAAGCGCTGCAGGACAAACCCCGGTTTCTCAACTTCGCGACCG  
AGCCGGGCGCACGCTGTACGGCGAGGACGTCTACGTCGGGTACAGGTACTACGAGTTTGCAGCAAGGACGTCAATTTC  
CCTTTGGCCACGGCCTGTCCTAACCACTTTTGCCCTTTTCCAATCTCTCCGTGTCTCACAGGACGGCAAGCTGAGCGTGT  
CCCTCTCCGTGAAGAACACCGGCTCCGTGCCGCGCACAGGTGGCCAGCTCTACGTCAAGCCCTCCAAGCGGCAAGA  
TTAACCGCCCGTCAAGGAGCTCAAGGGCTTCGCAAGGTGAACTGCAGCCCGCGAGACGAAAGGTGACTATGAGGTTATCG  
AGCAGGAGAGTACGTGCTGCGTATTTTGATGAGGAGCGGGATCAGTGGTGTGCGAAAAGGTGACTATGAGGTTATCG  
TGAGCGACAGCAGCGCAGCGAAGGATGGGTTGCGCTCAGGGTAAAGTTTACGGTGGGAGAGACGTATTGGTGGTCTGGCG  
TGTAAGTCTGTCATCATCTTTGGCAGATGAATCCAGTCACTTTGCATATAGCGCCGATGAAATGAGAACCAACGATCAT  
TGTGTAAAAA

Figure 2-A

Predicted *T. reesei* BGL4 amino acid sequence

MADIDVEAIL KKLTLAEKVD LLAGIDFWHT KALPKHGVP LRFDTGPNV RGTKFFNGVP AACFP CGTSL  
10 20 30 40 50 60 70

GSTFNQTLLE EAGKMMGKEA IAKSAHVILG PTINMQRSPL GGRGFESIGE DPFLAGLGAA ALIRGIQSTG  
80 90 100 110 120 130 140

VQATIKHFLC NDQEDRRMMV QSIVTERALR EIYALPFQIA VRDSQPGAFM TAYNGINGVS CSENPKYLDG  
150 160 170 180 190 200 210

MLRKEWGWDG LIMSDWYGY STTEAVVAGL DLEMPGPPRF RGETLKFNV NGKPFIHVID QRAREVLQFV  
220 230 240 250 260 270 280

KKCAASGVTE NGPETTVNNT PETAALLRKV GNEGIVLLKN ENNVLP LSKK KKT LIVGPNA KQATYHGGGS  
290 300 310 320 330 340 350

AALRAYYAVT PFDGLSKQLE TPPSYTVGAY TTVPPILGEQ CLTPDGAPGM RWRVFNEPPG TPNRQHIDEL  
360 370 380 390 400 410 420

FFTKTDMHLV DYYHPKAADT WYADMEGTYT ADEDCTYELG LVVCGTAKAY VDDQLVVDNA TKQVPGDAFF  
430 440 450 460 470 480 490

GSATREETGR INLVKGNTYK FKIEFGSAPT YTLKGDTIVP GHGSLRVGGC KVIDDQAEIE KSV ALAKEHD  
500 510 520 530 540 550 560

QVIICAGLNA DWETEGADRA SMKLPGLVDQ LIADVAAANP NTVVVMQTGT PEEMPWL DAT PAVIQAWYGG  
570 580 590 600 610 620 630

108121" 00042001

Figure 2-B

Predicted *T. reesei* BGL4 amino acid sequence

```
NETGNSIADV VFGDYNPSGK LSLSPKRLQ DNPAFLNFT EAGRTLYGED VYVGYYEYEF ADKDVNFPFG 640 650 660 670 680 690 700
HGLSYTTTFAF SNLSVSHKDG KLSVLSVKN TGSVPGAQVA QLYVKPLQAA KINRPVKELK GFAKVELQPG 710 720 730 740 750 760 770
ETKAVTIEEQ EKYVAAFFDE ERDQWCVEKG DYEIVSDSS AAKDGVALRG KFTVGETYWW SGV 780 790 800 810 820 830 833
```